Prediction of Total Phosphorus Using an Artificial Neural Network

Gebdang Biangbalbe Ruben, and Akoto Y. Boakye

Abstract—Rivers in urbanized cities and towns especially in developing countries experienced widespread eutrophication as a result of rapid dumping of waste which increase the phosphorus load in the environment and creates an imbalance in the ecosystem. Xuxi River is a typical case of eutrophic ecosystem with an abnormal growth of plants and death of the aquatic organisms. A simple and eco-friendly method using useful bacteria to degrade the pollutant load in the water were used in 2009. The purpose of this study is to estimate the total phosphorus by using artificial neural network (ANN) during the bacterial treatment. Seven water quality variables such as Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Total Nitrogen (TN), Suspended Sediment (SS), Temperature, Transparency, and Ammonia Nitrogen (NH₃-N) were applied as inputs for the network. The results showed 0.92 and 0.1031 for coefficient of correlation (r) and a root mean square error (RMSE) of respectively. To determine the influence of input variables on the dependent variable, a sensitivity analysis was carried out; the results revealed that DO, COD and NH₃-N are to be excluded from the input parameters because of their inaccuracy to predict TP. These results of this study demonstrate that the ANN model proposed can estimate the TP with a good degree of accuracy under bacteria technology.

Index Terms—Artificial neural network, prediction, total phosphorus.

I. INTRODUCTION

Lake Eutrophication is one of the most important water environment problems in China, which would lead to abundant development of aquatic plants, growth of algae, and disturb the balance of organisms in the water [1]. Due to the great migration of peoples from rural area to urban cities these last decades in China, water resources are under pressure. As a result of poor wastewater management from the community, urban rivers are subjected to sewage disposal and other waste. Many effort are being made to restore polluted water bodies, yet, lots have to be done in order to achieve the required water and sanitation demands. To meet these standards, many techniques have been employed over the years towards stream and river restoration as mentioned earlier. The most efficient method found recently is the Bacterial Technology, a technique that use bacteria to reduce pollutants in water. However, the use of such a method for the restoration of urbanized rivers in China is relatively new, it has been successfully used in the treatment of lakes, wastewater treatment plants and urban streams in Shenzhen, Rui’an and Wuxi of China. The method has received great concern because of the innovative and fast way it lowers the concentrations of the Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) concentrations [2]. This method could be more efficient if models for forecasting and monitoring of nutrient concentration during treatment campaigns are developed. Predictive modelling of nutrient storage in rivers is an important aid to monitoring exercises of urban rivers. As an illustration, Amos Kabo developed a mathematical model to predict the fate of BOD, NH₃-N and TP during the bacterial technology in various cities in China [3]-[5]. The main challenge remains the lack of sufficient data for the improvement of the statistical model’s accuracy. Therefore, models which does not require lots of data are of great importance for predicting total phosphorus. The estimation of total phosphorus presence in urban rivers is an important indicator for understanding eutrophication problems. So, this research developed an artificial neural network to model the fate of total phosphorus in bacteria treated urbanized rivers.

II. MATERIAL AND METHOD

A. Site and Datasets

The ecological river restoration took place in the Chang Nan District of Wuxi city of China. The river is characterized by muddy sediments sometimes up to 1.6m. The water quality of the river is very poor due to the daily sewage discharged into the River which was estimated to 10000m³ [2]. That makes the standard of surface water quality, to be ranked in class V. According to the Chinese National Standard (CNS) for Surface Water Quality (GB2828-2002), the Class V is the poorest water quality standard [6]. The Xuxi River then, needs to be restored to its natural condition for the benefit of the aquatic life and domestic uses.

The data used for this ANN were obtained from the bacterial technology in October 2009. The samples were collected at 8:30 AM to 4:30 PM during the experimental period and the physiochemical parameters of Temperature, dissolved oxygen (DO), chemical oxygen demand (COD), total nitrogen (TN), suspended sediment (SS), Transparency, and ammonia nitrogen (NH₃-N) were collected at five specific monitoring points of the Xuxi River.

B. The Bacterial Technology

The Bacterial Technology Method (BTM), system employs bacteria to degrade pollutants into simple harmless substances and produce standard effluents [7], provides a promising
advantage compared to the wetlands. It has been successfully implemented in the recovery of polluted lakes and various polluted systems such as wastewater treatment plants in Shenzhen, septic tanks in Menshenyuan District, and has also been used to restore an urban streams in Shenzhen, Rui’an and Wuxi, China [6]. This method is known to rapidly reduce the concentrations of effluent biological oxygen demand (BOD) and chemical oxygen demand (COD) [3]. It offers an easy solution to purify polluted streams and rivers and meets the standards of wastewater effluents without building massive structures as it is the case of traditional methods.

In fact, during River restoration campaigns, different technique - the re-aeration using a series of weir, shifting effluent discharge locations, pumping air into the water body using a local oxygenator and introducing a constructed wetland - have been used in China [6]. According to some studies, wetlands and their modifications are able to remove more than 70% of pollutants, which cost much less in construction, operation and maintenance than the conventional wastewater treatment plants [8]. However, the inability of constructed wetlands to efficiently remove phosphorus has been a major setback [9] that is why the bacterial technology is currently receiving more attention as an efficient, cost-effective and sustainable approach towards river restoration [3]. This method is simple, affordable, adaptable, scalable, and eco-friendly in river and wastewater treatment [10].

C. Artificial Neural Networks

An artificial neural network (ANN) is a mathematical structure designed to mimic the information processing functions of a network of neurons in the brain [11]. A feed forward Multilayer Perceptron (MLP) network, one of the various type of ANNs, consists of an input layer which receives the values of the input variables, an output layer which provides the model output, and one or more hidden layers. Nodes in each layer are interconnected through weighted acyclic arcs from each preceding layer to the following. It has been well recognized that a neural network with one hidden layer is capable of approximating any finite nonlinear function with high accuracy and three more hidden layered systems are known to cause unnecessary computational overload [12]. Hence, an MLP network with one hidden layer trained by backpropagation (BP) neural network was used to build the ANN model for modeling of the river water TP with seven input variables. Five to fifteen processing neurons and one single output which is TP were applied. The ANNs modeling strategy is implemented by Neural Network Toolbox in MATLAB®. The neural architecture involves three different layer: (i) an input layer, (ii) a hidden layer, (iii) an output layer. The number of neurons in the hidden layers is determined by trial-and-error, and the neurons of each layer is connected to the neurons of the next layer by weights. The network properties are as follows:

- Network inputs: DO, Transparency, COD, TN, NH₃-N, SS, Temperature
- Network output: TP
- Network type: Feed-Forward Back-Propagation

- Training function: Levenberg–Marquardt (TRAINLM)
- Activation function (hidden layer): tan-sigmoid
- Activation function (output layer): linear function (PURELIN)
- Performance function: MSE
- Data division: dividerand.

D. Data Processing

The parameters analyzed in this paper are based on variables that reflect the water quality: stratification (temperature, suspended sediment, transparency, and dissolved oxygen), the trophic status (total nitrogen, and ammonia-nitrogen), and the organic matter content (chemical oxygen demand). To build the Artificial Neural Network, the data have to be divided into training, validation and testing datasets. The training set is used to estimate the unknown connection weights; the validation set is used to decide when to stop training in order to avoid overfitting and/or which network structure is optimal; and the test set is used to assess the generalization ability of the trained model [13]. Overall, 110 data samples collected from the river were divided into two groups as training (90 samples) and testing using 20 randomly selected data. The testing data set was an independent validation data set completely separated from the first one in order to test the predictability of the network or in other word to evaluate the optimized model against unknown data set.

III. RESULTS AND DISCUSSION

A. TP model results

In this study, performance of the models is assessed in accordance with the root mean square error (RMSE) and the coefficient of correlation (r) between the observed and the predicted values. The total phosphorus (TP) was modelled as an output variable against seven input parameters (TN, Transparency, Temperature, NH₃-N, SS, COD and DO). Many iterations of the MLP were performed with the change of the number of neurons from 5 to 15. The best number of neurons were determined by the minimum value of MSE of the test data set and a higher coefficient of correlation(r). The results obtained were evaluated with an independent data set for their predictive power and reliability for practical forecasting under the BTM campaigns. The results obtained with 10 neurons have shown to be the best, which gives an architecture of 7-10-1 with 7 inputs parameters, 10 hidden neurons, and 1 output representing TP.

Fig.1 depicts the observed TP removals against the corresponding measured TP removals. The coefficient of correlation gives, r = 0.92, the root mean square error (RMSE) equal to 0.1, and the mean square error (MSE) of 0.22. The results imply that the neural networks can predict the TP removal under the bacterial technology due to the strong correlation between the observed values and the predicted ones.

In order to check the influence of each parameter on the end result and TP, a sensitivity analysis needs to be carried out.
B. Sensitivity Analysis

To evaluate the effect of inputs variables on the TP model, we evaluated the performance of various combinations of the parameters by using the coefficient of correlation (r), and the RMSE approaches. These produced various possible models that could be used to practically estimate the values of TP. The optimal network architecture of the various combinations of the parameters was selected based on the one with minimum MSE using the 10 neurons. Overall, seven networks were compared as shown in Table I. The results show that by eliminating some variables, the accuracy of the prediction improve while others decline drastically. We observed that by removing the DO, the precision of model is highly accurate where minimum RMSE and coefficient of correlation (r) were determined to be 0.4304 and 0.9579 for the test data set respectively. The same observation has been made for the NH$_3$-N whose correlation coefficient and RMSE give 0.9842 and 0.0454 respectively. By eliminating the COD from the input variable too, the results show 0.9665 and 1.0022 for correlation coefficient and RMSE correspondingly. Therefore, from these results, the DO, NH$_3$-N and COD could be excluded, they are not relevant to the TP model. Fig.2 and Fig.3 illustrate this deduction better, where the observed TP value and predicted TP value are closely matching. All plots show a relatively high correlation between the observed and predicted values. The 3 models (eliminate DO, eliminate NH$_3$-N, eliminate COD) derived here are therefore acceptable, could then be used for TP prediction and can be applied to support decision making but are not explicit in replacing traditional field measurements. Based on the results above, for the estimation of total phosphorus during the bacteria technology with an artificial neural network, the following parameters should be assessed in situ carefully: total nitrogen, temperature, transparency, and suspended sediment.

<table>
<thead>
<tr>
<th>Model</th>
<th>Combination</th>
<th>Structure</th>
<th>RMSE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All</td>
<td>7-10-1</td>
<td>0.1031</td>
<td>0.9246</td>
</tr>
<tr>
<td>2</td>
<td>Eliminate DO</td>
<td>6-10-1</td>
<td>0.4304</td>
<td>0.9579</td>
</tr>
<tr>
<td>3</td>
<td>Eliminate COD</td>
<td>6-10-1</td>
<td>0.2242</td>
<td>0.7207</td>
</tr>
<tr>
<td>4</td>
<td>Eliminate TN</td>
<td>6-10-1</td>
<td>1.0022</td>
<td>0.9665</td>
</tr>
<tr>
<td>5</td>
<td>Eliminate Temperature</td>
<td>6-10-1</td>
<td>2.2499</td>
<td>0.659</td>
</tr>
<tr>
<td>6</td>
<td>Eliminate Transparency</td>
<td>6-10-1</td>
<td>0.1425</td>
<td>0.8701</td>
</tr>
<tr>
<td>7</td>
<td>Eliminate Suspended Sediment</td>
<td>6-10-1</td>
<td>0.1395</td>
<td>0.9229</td>
</tr>
<tr>
<td>8</td>
<td>Eliminate NH$_3$-N</td>
<td>6-10-1</td>
<td>0.0454</td>
<td>0.9842</td>
</tr>
</tbody>
</table>

Fig.1 Comparison of observed TP value and predicted TP value (left); correlation between observed and predicted TP (right).
IV. CONCLUSION

During River restoration campaigns using an eco-friendly method called the bacterial technology, contaminants including the total phosphorus (TP) need to be measured regularly several times a day which require both time and labor. A solution to this, is artificial neural network made of one input layer, one hidden layer and one output layer trained with the Levenberg–Marquard algorithm was used to estimate the total phosphorus. By a trial and error method, the best structure of the networks were selected. The final structure 7 - 10 - 1, was made up of seven water quality variables as the inputs (DO, COD, TN, SS, Temperature, Transparency, and NH$_3$-N) in order to obtain the TP as the output. Based on the minimum value of MSE and the higher value of coefficient of correlation, of the training data set, the best network after training showed a coefficient of correlation ($r$) and a root mean square error (RMSE) of 0.92 and 0.1031 respectively, with a good fit between the observed and the predicted TP. A sensitivity analysis was carried out in order to evaluate the effects of the seven inputs variables on the TP model. The results show that DO, COD and NH$_3$-N are to be excluded because of their inaccuracy to predict TP. By removing the DO from the input parameters, for example, the precision of prediction get better with RMSE and coefficient of correlation ($r$) of 0.4304 and 0.9579 respectively for the data set testing. Four input variables (TN, Temperature, Transparency, and SS) affect TP significantly, thus could be used for further prediction for a...
more efficient result. These results demonstrate that the ANN proposed in this study can estimate the TP with a good degree of accuracy under bacteria technology. The derived models in this research envisage a contribution towards water quality monitoring and management of urban rivers and streams in China. It is also believed that, urban rivers and streams in other parts of the world with similar conditions of the class V Chinese national board standard grouping, might be applicable to use the models developed here.

REFERENCES


Gebdang B. Ruben, born on 6th June, 1989 at Bertoua, Cameroon. Currently, Ruben is a Master’s student at Hohai University, China, majoring in Hydrology and water resource. The degree will be issued in June 2016. In 2012, Mr. Gebdang earned a Bachelor degree at the University of Dschang, Cameroon in the field of Water supply engineering. From June – August 2015, he did an internship at the Laboratory of Water Pollution Control and Soil Remediation. He was in charge of analyzing polluted water from Xiaoliu River, Nanjing, China. In 2012, he did four months (May 2012-September 2012) of internship at Hydrofor, a company that supplies drinking water by drilling in the North-Cameroon. One year before, he was a TRAINEE at the Regional Delegation of Water an Energy for the North Region of Cameroon, where he spent two months (July 2011-September 2011). He has a paper to be published titled “Estimation of Ammonia-Nitrogen (NH₃-N) using an artificial neural network under the bacterial technology”, Oriental Journal of Chemistry, 2015.

Yiadom Boakye Akoto, born on 12th November, 1985 at Kumasi, Ghana. Currently, Akoto is pursuing Master’s program in coastal engineering at the Hohai University, China. The degree will be issued in June 2017. Mr. Akoto had his BSc. degree in civil engineering at the University of Science and Technology, Kumasi, Ghana.

He is now working with the Hydrological services department, of MWRWH, Accra, Ghana.